CLAIMS

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

1. A method of processing an inductive learning model for a dataset of examples, said method comprising:

dividing said dataset into a plurality of subsets of data; and

developing an estimated learning model for said dataset by developing a learning model for a first subset of said plurality of subsets.

2. The method of claim 1, further comprising:

progressively forming an ensemble model of said dataset by sequentially developing a learning model for each of a successive one of said plurality of subsets, until a desired indication of termination has been reached.

3. The method of claim 1, further comprising:

developing at least one of a current accuracy and an estimated final accuracy,

said current accuracy comprising an accuracy of said learning model for said first subset,

said estimated final accuracy comprising an estimated accuracy of said estimated learning model for said dataset.

4. The method of claim 2, further comprising:

developing at least one of a current accuracy and an estimated final accuracy,

said current accuracy comprising an accuracy of said learning model for said subset being currently developed,

said estimated final accuracy comprising an estimated accuracy of said ensemble model of said dataset.

5. The method of claim 2, further comprising:

developing an estimated training time to complete development of said ensemble model.

6. The method of claim 3, wherein each said example in said dataset carries a benefit and said accuracy comprises an overall accuracy that reflects an estimated total amount of reward from said benefits.

7. The method of claim 6, wherein said benefit is not equal for all said examples, said learning comprising a cost-sensitive learning, and said accuracy comprises an overall accuracy that reflects an estimated total amount of reward from said benefits in units of money.

8. An apparatus for processing an inductive learning model for a dataset of examples, said apparatus comprising:

a database divider for dividing said dataset into N subsets of data; and

a base classifier calculator for developing a learning model for data in a first subset of said N subsets.

9. The apparatus of claim 8, further comprising:

an ensemble calculator for progressively developing an ensemble model of said database of examples by successively integrating a base classifier from successive subsets of said N subsets.

10. The apparatus of claim 9, further comprising:

a memory interface to retrieve data from said database and to store data as said inductive learning model is progressively developed; and

a graphic user interface to allow a user to selectively enter parameters, to control the progressive development of said ensemble model, and to view results of said progressive development.

11. A system to process an inductive learning model for a dataset of example data, said system comprising one or more of:

a memory containing one or more of a plurality of segments of said example data, wherein each said segment of example data comprises data for calculating a base classifier for an ensemble model of said dataset;

a base classifier calculator for developing a learning model for data in one of said N segments;

an ensemble calculator for progressively developing an ensemble model of said database of examples by successively integrating a base classifier from successive ones of said N segments;

a memory interface to retrieve data from said database and to store data as said inductive learning model is progressively developed; and

a graphic user interface to allow a user to at least one of enter parameters, to control the progressive development of said ensemble model, and at least one of display and printout results of said progressive development.

12. A method of providing a service, said method comprising at least one of:

providing a database of example data to be used to process an inductive learning model for said example data, wherein said inductive learning model is derivable by dividing said example data into N segments and using at least one of said N segments of example data to derive a base classifier model;

receiving said database of example data and executing said method of deriving said inductive learning model;

providing an inductive learning model as derived;

executing an application of an inductive learning model as derived; and

receiving a result of said executing said application.

13. A method of deploying computing infrastructure, comprising integrating computer-readable code into a computing system, wherein the code in combination with the computing system is capable of processing an inductive learning model for a dataset of examples by:

dividing said dataset into N subsets of data; and

developing an estimated learning model for said dataset by developing a learning model for a first subset of said N subsets.

14. A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method of processing an inductive learning model for a dataset of examples, said method comprising:

dividing said dataset into N subsets of data; and

developing an estimated learning model for said dataset by developing a learning model for a first subset of said N subsets.

15. The signal-bearing medium of claim 14, said method further comprising:

progressively forming an ensemble model of said dataset by sequentially developing a learning model for each of a successive one of said N subsets, until a desired indication of termination has been reached.

16. The signal-bearing medium of claim 15, said method further comprising:

developing at least one of a current accuracy and an estimated final accuracy,

said current accuracy comprising an accuracy of said learning model for said subset being currently developed,

said estimated final accuracy comprising an estimated accuracy of said ensemble model of said dataset.

17. The signal-bearing medium of claim 15, said method further comprising:

developing an estimated training time to complete development of said ensemble model.

18. The signal-bearing medium of claim 16, wherein each said example in said dataset carries a benefit and said accuracy comprises an overall accuracy that reflects an estimated total amount of reward from said benefits.

19. The signal-bearing medium of claim 18, wherein said benefit is not equal for all said examples, said learning comprising a cost-sensitive learning, and said accuracy comprises an overall accuracy that reflects an estimated total amount of reward from said benefits in predetermined units.

20. A method of at least one of increasing a speed of development of a learning model for a dataset of examples and increasing an accuracy of said learning model, said method comprising:

dividing said dataset into N subsets of data; and

developing an estimated learning model for said dataset by developing a learning model for a first subset of said N subsets.

21. The method of claim 20, further comprising:

calculating an estimated accuracy for said learning model.

22. The method of claim 20, further comprising:

calculating a remaining training time.

23. The method of claim 20, further comprising:

progressively, and stepwise, forming an ensemble model of said dataset by sequentially

using additional said subsets to develop an additional learning model for said subset and incorporating each said additional learning model into an aggregate model to form said ensemble model, wherein said progressive and stepwise forming can be terminated prior to developing an additional learning model for all of said N subsets.

24. The method of claim 20, wherein said examples carry potentially different benefits, said method further comprising:

calculating an estimation of an accumulated benefit for said learning model.

25. A method of developing a predictive model, said method comprising:

for a dataset comprising a plurality of elements, each said element comprising a feature vector, said dataset further comprising a true class label for at least a portion of said plurality of elements, said true class labels allowing said dataset to be characterized as having a plurality of classes, dividing at least a part of said portion of said plurality of elements having said true class label into N segments of elements; and

learning a model for elements in at least one of said N segments, as an estimate for a model for all of said dataset.

26. The method of claim 25, further comprising:

using a second part of said portion of said plurality of elements having said true class label as a validation set for said model.

27. The method of claim 26, further comprising:

using said validation set to calculate a predicted accuracy for said model.

28. The method of claim 25, further comprising:

calculating an estimated training time for learning a model based on a remainder of said N segments.

29. The method of claim 25, wherein said elements are each associated with a benefit, said method further comprising:

establishing a benefit matrix associated with said plurality of classes, said benefit matrix defining a benefit for each said element in said dataset as applicable for each said class.

30. The method of claim 29, wherein said elements in said dataset can respectively have different benefit values, said method further comprising:

using a validation dataset to measure a validation of said model; and

calculating an aggregate benefit for said model, as based on said validation dataset.

31. The method of claim 25, further comprising:

progressively developing an ensemble model by successively learning a model for elements in one of a remaining said N segments, wherein said progressively developing said ensemble model is terminable at any stage.

32. The method of claim 31, further comprising:

calculating at least one of an accuracy and a remaining training time for said ensemble

model.

33. The method of claim 32, further comprising:

entering a threshold for at least one of said accuracy and said remaining training time; and

automatically terminating said progressively developing said ensemble model whenever said threshold is exceeded.